CONGESTION CONTROL IN A WIRELESS DATA NETWORK

CLAIM OF PRIORITY UNDER 35 U.S.C. SECTION 119

The present application is a continuation of U.S. application Ser. No. 10/646,242, filed Aug. 22, 2003, entitled "CONGESTION CONTROL IN A WIRELESS DATA NETWORK," which claims priority to provisional application Ser. No. 60/448,269, entitled "REVERSE LINK DATA COMMUNICATION," filed on Feb. 18, 2003; U.S. provisional application Ser. No. 60/452,790, entitled "METHOD AND APPARATUS FOR A REVERSE LINK COMMUNICATION IN A COMMUNICATION SYSTEM," filed on Mar. 6, 2003; U.S. provisional application, and Ser. No. 60/470,770, entitled "OUTER-LOOP POWER CONTROL FOR REL. D," filed on May 14, 2003.

FIELD

The present invention relates generally to wireless communications, and more specifically to a novel and improved method and apparatus for congestion control in a wireless data network.

BACKGROUND

Wireless communication systems are widely deployed to provide various types of communication such as voice and data. These systems may be based on code division multiple access (CDMA), time division multiple access (TDMA), or some other multiple access techniques. A CDMA system provides certain advantages over other types of systems, including increased system capacity.

A CDMA system may be designed to support one or more CDMA standards such as (1) the "TIA/EIA-95-B Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System" (the IS-95 standard), (2) the standard offered by a consortium named "3rd Generation Partnership Project" (3GPP) and embodied in a set of documents including Document Nos. 3G TS 25.211, 3G TS 25.212, 3G TS 25.213, and 3G TS 25.214 (the W-CDMA standard), (3) the standard offered by a consortium named "3rd Generation Partnership Project 2" (3GPP2) and embodied in "TR-45.5 Physical Layer Standard for cdma2000 Spread Spectrum Systems" (the IS-2000 standard), and (4) some other standards.

In the above named standards, the available spectrum is shared simultaneously among a number of users, and techniques such as power control and soft handoff are employed to maintain sufficient quality to support delay-sensitive services, such as voice. Data services are also available. More recently, systems have been proposed that enhance the capacity for data services by using higher order modulation, very fast feedback of Carrier to Interference ratio (C/I) from the mobile station, very fast scheduling, and scheduling for services that have more relaxed delay requirements. An example of such a data-only communication system using these techniques is the high data rate (HDR) system that conforms to the TIA/EIA/IS-856 standard (the IS-856 standard).

In contrast to the other above named standards, an IS-856 system uses the entire spectrum available in each cell to transmit data to a single user at one time, selected based on link quality. In so doing, the system spends a greater percentage of time sending data at higher rates when the channel is good, and thereby reduces committing resources to support transmission at inefficient rates. The net effect is higher data capacity, higher peak data rates, and higher average system throughput.

Systems can incorporate support for delay-sensitive data, such as voice channels or data channels supported in the IS-2000 standard, along with support for packet data services such as those described in the IS-856 standard. The cdma2000® Revision C of the IS-2000 standard (including C.S0001.C through C.S0006.C) is such a system, and is hereinafter referred to as the 1xEV-DV system. In the rest of the document, we'll refer to release 0, A, and B of the cdma2000® standard as cdma2000, while revision C and upwards will be referred to as 1xEV-DV systems.

An example 1xEV-DV system includes a reverse link control mechanism for allocating the shared reverse link resource for transmission by a plurality of mobile stations. A mobile station may make a request to a serving base station for transmission permission with a maximum rate supportable by the mobile station. Alternatively, a mobile station is allowed to transmit autonomously, without making a request, at a rate up to a determined autonomous maximum rate. The serving base station anticipates an expected amount of autonomous transmission on the reverse link, reviews any requests made by mobile stations, and allocates the shared resource accordingly. The base station may elect to make one or more individual grants to requesting mobile stations, and includes the maximum rate for those grants. The remaining requesting mobile stations may be issued permission to transmit according to a common grant, with an associated maximum transmission rate. Thus, the serving base station attempts to maximize utilization of the shared resource with a combination of individual and common grants, in the presence of autonomous transmission by other mobile stations. Various techniques may be used to allow mobile stations to continue transmitting according to a determined allocation and the associated grants, with a minimum amount of signaling required.

From time to time, the amount of loading on the reverse link may exceed the amount predicted by the serving base station. Various factors may lead to this system over-utilization, an example of which is the uncertainty in the actual number of autonomous transmissions that may transpire. Overall throughput, and thus effective capacity of the system, may deteriorate when the system becomes congested. For example, a resultant increase in error rate may result in loss of successful data transmission, and subsequent retransmission will use additional capacity on the shared resource. While the allocation and granting procedure just described may be used to alleviate overloading on the system, there is latency associated with the messaging required. Capacity and throughput may be adversely affected during this time. It is desirable to be able to reduce the system loading quickly to minimize these adverse effects.

Furthermore, additional messaging also uses system capacity. In some circumstances, the system overload is a temporal condition, after which, the previous allocation and associated grants will be appropriate for the desired system load. It is desirable for the various mobile stations to return to the prescribed allocation while minimizing messaging overhead. There is therefore a need in the art for congestion control to reduce system loading efficiently.

SUMMARY

Embodiments disclosed herein address the need for congestion control. In one embodiment, a base station allocates a shared resource using a combination of zero or more indi-